Turkiye Klinikleri J Biostat 2018;10(1):1-12

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2017-59265

A Discrete Akash Distribution with Applications

Kesikli Akash Dağılımı ve Uygulamaları

Berhane ABEBE,[a]
Rama SHANKER[a]

[a]Department of Statistics,
College of Science,
Eritrea Institute of Technology,
Asmara, Eritrea

Received: 18.12.2017
Received in revised form: 30.01.2018
Accepted: 05.02.2018
Available online: 05.04.2018

Correspondence:
Rama SHANKER
College of Science,
Eritrea Institute of Technology,
Department of Statistics, Asmara,
ERITREA/ERİTRE
shankerrama2009@gmail.com

ABSTRACT The discretization of continuous distribution has drawn the attention of researchers in the recent decades due to the fact that it generates distributions that can be used for strictly discrete data. In this paper, a discrete Akash distribution, a discrete analogue of continuous Akash distribution, has been proposed and investigated. Its moment generating function, moments and moments based measures including coefficients of variation, skewness, kurtosis and index of dispersion have been obtained and discussed. For estimating its parameter the method of moments and the method of maximum likelihood estimation have been discussed. The usefulness and the goodness of fit of the proposed distribution have been explained using some real datasets and found that it gives better fit as compared to other discrete distributions.

Keywords: Akash distribution; discretization; moment generating function; moments; estimation; goodness of fit

ÖZET Son yıllarda tamamen kesikli veri için kullanılabilen dağılımlar oluşturulması nedeniyle sürekli bir dağılımın kesikli hale getirilmesi araştırmacıların ilgisini çekmiştir. Bu makalede sürekli Akash dağılımının kesikli benzeri olan, kesikli Akash dağılımı sunulmuş ve araştırılmıştır. Dağılımın moment çıkaran fonksiyonu, momentleri ve momentlere dayalı olan değişim katsayısı, çarpıklık, basıklık ve yayılım indeksi de elde edilmiş ve tartışılmıştır. Parametrelerin tahmini için moment ve en çok olabilirlik yöntemi kullanılmıştır. Önerilen dağılımın uyum iyiliği ve kullanışlılığı bazı gerçek veri setleriyle açıklanmış ve diğer kesikli dağılımlarla karşılaştırıldığında daha iyi sonuçlar vermiştir.

Anahtar Kelimeler: Akash dağılımı; kesikli hale getirme; moment çıkaran fonksiyon; momentler; tahmin; uyum iyiliği

In recent decades, the discretization of continuous distributions has been widely considered. The main reasons for discretizing a continuous distributions are twofold, namely, (i) the discrete analogue of a continuous distribution provide probability mass function (pmf) that can compete with the classical discrete distributions commonly used in the analysis of count data and (ii) the discrete analogue of a continuous

Copyright © 2018 by Türkiye Klinikleri

distribution avoids the use of a continuous distribution in the case of strictly discrete data.

In many real life situations, it is difficult or inconvenient to get samples from continuous distributions. The observed values, in general and almost always, are actually discrete in nature because they are measured to only finite number of decimal places and cannot really constitute all points in a continuum. In case of lifetime data (waiting time or survival time), even if the measurements are taken on a continuous scale, the observations may be recorded in a way that makes a discrete distribution more appropriate model. For example, in survival analysis, it is most common to use continuous distributions to model discrete data. A discretization of a continuous distribution acts as a subterfuge to avoid the use of continuous distribution to model survival time data. According to Lai (2013), discretization of a continuous lifetime model is an interesting and intuitively appealing approach to derive a discrete lifetime model corresponding to the continuous one.[1] It has been observed that in real world the original variables may be continuous in nature but discrete by observation and, therefore, it is reasonable and convenient to model the situation by an appropriate discrete distribution generated from the underlying continuous distribution preserving one or more important characteristics including probability density function (pdf), moment generating function (mgf), moments, hazard rate function, mean residual life function etc., of the continuous distribution.

There are several methods available in Statistics literature to derive a discrete distribution from a continuous distribution. One of the first proposed discretization methods is based on the definition of pmf that depends on an infinite series. The method of discretization by an infinite series was firstly considered by Good (1953) who has proposed the discrete Good distribution to model the population frequencies of species and the estimation of parameters.[2] A random variable Y is said to have a discrete Good distribution if its pmf can be expressed as

$$P(Y=y)=\frac{\alpha^{y} y^{\beta}}{\sum_{j=0}^{\infty}\alpha^{j} j^{\beta}} \;;\; y=0,1,2,.... \tag{1.1}$$

where $\beta \in R$ and $\alpha \in (0,1)$.

The method of infinite series is characterized by the following definition.

Definition 1.1: Let X be a continuous random variable having pdf $f_X(x)$ with support on R. Then the corresponding discrete random variable Y has pmf given by

$$P(Y=y)=P(y;\theta)=\frac{f_X(y;\theta)}{\sum_{j=-\infty}^{\infty} f_X(j;\theta)} \;;\; y\in Z \;, \tag{1.2}$$

where θ may be the vector of parameters indexing the distribution of X .

This method of discretizing a continuous distribution has been studied by several researchers including Kulasekara and Tonkyn (1992), Doray and Luong (1997), Sato *et al* (1999), Nekoukhou *et al* (2012), are some among others, who proposed a version of the method when the continuous random variable of interest is defined on R_+.[3-6] Thus, if the random variable X is defined on R_+, the pmf of Y can be defined as

$$P(Y=y)=P(y;\theta)=\frac{f_X(y;\theta)}{\sum_{j=0}^{\infty} f_X(j;\theta)} \quad ; y\in Z_+ \tag{1.3}$$

Recently Josmar *et al* (2017), using above method of discretization, has obtained a discrete Shanker distribution (DSD) with parameter $\theta>0$ and having pmf[7]

$$P_1(y;\theta)=\frac{\left(e^{\theta}-1\right)^2(\theta+y)e^{-\theta(y+1)}}{1+\left(e^{\theta}-1\right)\theta};\ y=0,1,2,\ldots\ . \tag{1.4}$$

Josmar *et al* (2017) have discussed its statistical properties, estimation of parameter and applications to model count datasets.[7] Recall that DSD is a discrete analogue of continuous Shanker distribution proposed by Shanker (2015 a) having pdf[8]

$$f_1(x;\theta)=\frac{\theta^2}{\theta^2+1}(\theta+x)e^{-\theta x} \quad ;x>0,\ \theta>0 \tag{1.5}$$

Using infinite series method of discretization, the pmf of discrete Lindley distribution (DLD) obtained by Berhane and Shanker (2017) is given by [9]

$$P_2(y;\theta)=\frac{\left(e^{\theta}-1\right)^2}{e^{2\theta}}(1+y)e^{-\theta y};\ y=0,1,2,\ldots \tag{1.6}$$

where the parameter $\theta>0$.

Recall that the DLD is a discrete analogue of continuous Lindley distribution introduced by Lindley (1958) having pdf[10]

$$f_2(x;\theta)=\frac{\theta^2}{\theta+1}(1+x)e^{-\theta x} \quad ;x>0,\ \theta>0. \tag{1.7}$$

Note that the discrete distribution generated using the method of infinite series may not always have a compact form due to the normalizing constant.

A second common method of discretization of a continuous distribution is based on the survival function of the original continuous distribution and was proposed by Nakagawa and Osaki (1975) and has the interesting feature of preserving the original survival function on its integer part for the generated pmf [11]. Kemp (2004) pointed that a discrete random variable Y from a continuous random variable X can be defined as follows:[12]

Definition 1.2: Let X be a continuous random variable having survival function $S_X(x)=1-F_X(x)=P(X\geq x)$. Then , the discrete random variable $Y=\lfloor X\rfloor$ has pmf given by

$$P(Y=y)=P(y;\theta)=S_X(y;\theta)-S_X(y+1;\theta);\ y=0,1,2,\ldots\ , \tag{1.8}$$

where $Y=\lfloor X \rfloor$ = largest integer less than or equal to X .

Note that the resulting pmf will be in closed form if the original survival function has closed form. Using the method of survival function of distribution, Nakagawa and Osaki (1975) proposed a discrete Weibull distribution, a discrete analogue of Weibull distribution, and studied its properties, estimation of parameters and applications.[12]

In the present paper, a discrete Akash distribution, a discrete analogue of continuous Akash distribution proposed by Shanker (2015 b) has been introduced using a discretization method based on an infinite series.[13] Its moment generating function, moments and moments based measures have been obtained and discussed. The estimation of parameter of the distribution has been discussed using both the method of moments and the method of maximum likelihood estimation. The usefulness and the goodness of fit of the proposed distribution have been explained using some real datasets and the fit has been compared with other discrete distributions.

A DISCRETE AKASH DISTRIBUTION

The pdf and the cdf of a continuous random variable X having Akash distribution introduced by Shanker (2015 b) are given by[13]

$$f_2(x;\theta)=\frac{\theta^3}{\theta^2+2}\left(1+x^2\right)e^{-\theta x};\quad x>0,\quad \theta>0 \tag{2.1}$$

$$F_1(x;\theta)=1-\left[1+\frac{\theta x(\theta x+2)}{\theta^2+2}\right]e^{-\theta x}\quad ;x>0,\theta>0 \tag{2.2}$$

Shanker (2015 b) has discussed its various mathematical and statistical properties including its shapes for varying values of parameters, moments based measures, hazard rate function, mean residual life function, stochastic ordering, mean deviations, order statistics, Renyi entropy measure, Bonferroni and Lorenz curves and stress-strength reliability along with estimation of parameter and applications of the distribution for modeling lifetime data from biomedical sciences and engineering.[13] Shanker (2017) has also introduced a Poisson-Akash distribution (PAD), a Poisson mixture of Akash distribution having pmf[14]

$$P_3(x;\theta)=\frac{\theta^3}{\theta^2+2}\cdot\frac{x^2+3x+\left(\theta^2+2\theta+3\right)}{(\theta+1)^{x+3}}\ ;x=0,1,2,\ldots,\theta>0\ . \tag{2.3}$$

Shanker (2017) has discussed its various statistical and mathematical properties, estimation of parameter using both the method of moments and the method of maximum likelihood estimation and applications to model count data.[14]

Using the definition (1.1), the pmf of the discrete random variable Y, corresponding to a continuous random variable X following Akash distribution (2.1) with parameter $\theta>0$, can be obtained as

$$P_4(y;\theta)=\frac{\left(e^{\theta}-1\right)^3}{e^{\theta}\left(e^{2\theta}-e^{\theta}+2\right)}\left(1+y^2\right)e^{-\theta y};\ y=0,1,2,\ldots. \tag{2.4}$$

We would call this distribution, a discrete Akash distribution (DAD). The nature and behavior of DAD for varying values of its parameter θ has been shown graphically in Figure 1.

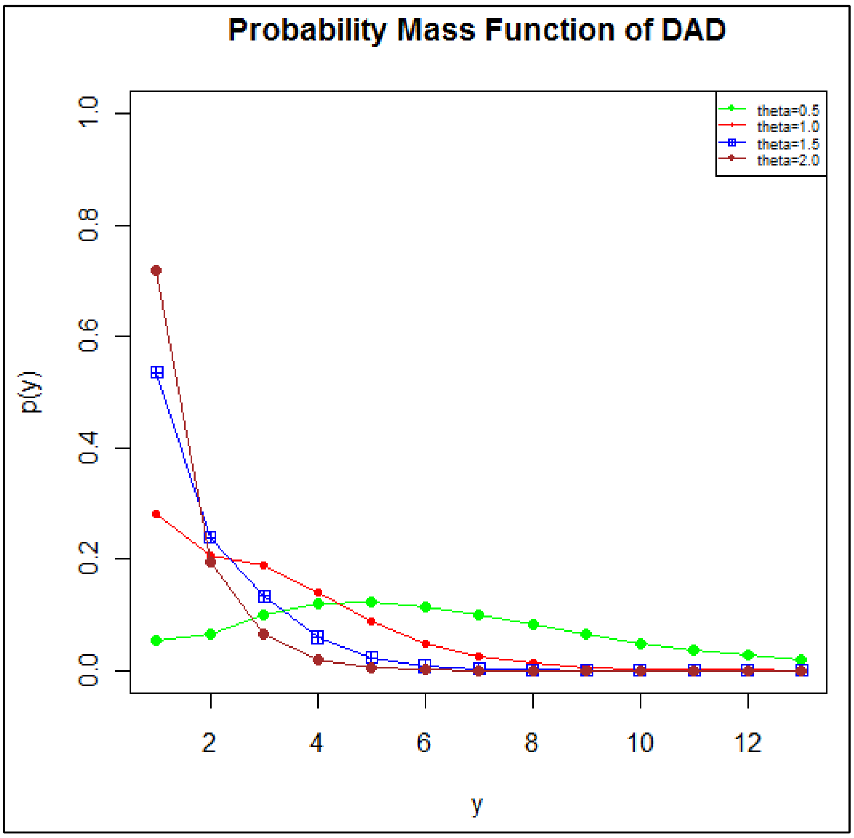

Figure 1: The pmf plot of DAD for varying values of the parameter θ.

The survival function, $S(y;\theta)$ and the cumulative distribution function (cdf), $F(y;\theta)$ of DAD can be obtained as

$$S(y;\theta)=\left[\frac{\left\{\left(e^{\theta}-1\right)y+1\right\}^{2}+e^{\theta}\left(2e^{\theta}-1\right)}{e^{\theta}\left(e^{2\theta}-e^{\theta}+2\right)}\right]e^{-\theta y};\, y=0,1,2,...,\theta>0 \tag{2.5}$$

$$F_2(y;\theta)=1-\left[\frac{\left\{\left(e^{\theta}-1\right)y+1\right\}^{2}+e^{\theta}\left(2e^{\theta}-1\right)}{e^{\theta}\left(e^{2\theta}-e^{\theta}+2\right)}\right]e^{-\theta y};\, y=0,1,2,...,\theta>0 \tag{2.6}$$

Graph of cumulative distribution function of DAD has been shown in Figure 2.

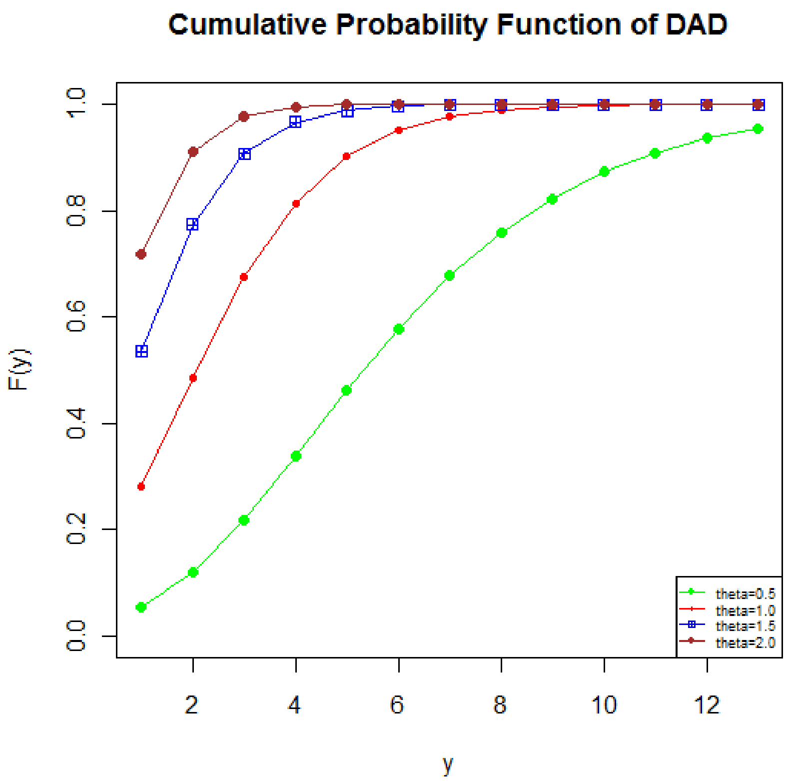

FIGURE 2: The cdf plot of DAD for varying values of the parameter θ.

Since $\frac{P(y+1;\theta)}{P(y;\theta)}=\left[1+\frac{2y+1}{1+y^2}\right]e^{-\theta}$ is a decreasing function of $y\geq 3$, $P(y;\theta)$ is log-concave and therefore, the DAD has an increasing hazard rate. Further, $\left[P(y;\theta)\right]^2\geq P(y-1;\theta)\cdot P(y+1;\theta)$ for $y\geq 2$, which implies unimodality, by theorem 3 of Keilson and Gerber (1971)[15]. The interrelationship between log-concavity, unimodality and increasing hazard rate of discrete distributions are available in Grandell (1997).[16]

MOMENTS AND ASSOCIATED MEASURES

The probability generating function (pgf) and the moment generating function (mgf) of DAD can be obtained as

$$G(t)=\frac{\left(e^{\theta}-1\right)^3}{e^{\theta}\left(e^{2\theta}-e^{\theta}+2\right)}\left[\frac{e^{\theta}\left(e^{\theta}-t\right)^2+t\,e^{\theta}\left(e^{\theta}-t\right)}{\left(e^{\theta}-t\right)^3}\right],\text{for } t\neq e^{\theta}, \tag{3.1}$$

and

$$M(t)=\frac{\left(e^{\theta}-1\right)^3}{e^{\theta}\left(e^{2\theta}-e^{\theta}+2\right)}\left[\frac{e^{\theta}\left(1-e^{-(\theta-t)}\right)^2+e^{-(\theta-t)}\left(1+e^{-(\theta-t)}\right)}{\left(1-e^{-(\theta-t)}\right)^3}\right],\text{for } t\neq\theta. \tag{3.2}$$

It can be easily verified that the function in (3.2) is infinitely differentiable with respect to t, since it involves exponential terms of its argument. This means that one can derive all moments about origin μ_r' , $r \geq 1$ of DAD from its mgf.

The first four moments about origin of DAD can thus be obtained as

$$\mu_1' = \frac{2\left(e^{2\theta}+e^{\theta}+1\right)}{\left(e^{2\theta}-e^{\theta}+2\right)\left(e^{\theta}-1\right)}$$

$$\mu_2' = \frac{2\left(e^{3\theta}+5e^{2\theta}+5e^{\theta}+1\right)}{\left(e^{2\theta}-e^{\theta}+2\right)\left(e^{\theta}-1\right)^2}$$

$$\mu_3' = \frac{2\left(e^{4\theta}+14e^{3\theta}+30e^{2\theta}+14e^{\theta}+1\right)}{\left(e^{2\theta}-e^{\theta}+2\right)\left(e^{\theta}-1\right)^3}$$

$$\mu_4' = \frac{2\left(e^{5\theta}+33e^{4\theta}+146e^{3\theta}+146e^{2\theta}+33e^{\theta}+1\right)}{\left(e^{2\theta}-e^{\theta}+2\right)\left(e^{\theta}-1\right)^4}$$

Using the relationship $\mu_r = E\left(Y-\mu_1'\right)^r = \sum_{k=0}^{r}\binom{r}{k}\mu_k'\left(-\mu_1'\right)^{r-k}$ between central moments and moments about origin, the central moments of DAD are obtained as

$$\mu_2 = \frac{2e^{\theta}\left(e^{4\theta}+2e^{3\theta}-2e^{2\theta}+5\right)}{\left(e^{2\theta}-e^{\theta}+2\right)^2\left(e^{\theta}-1\right)^2}$$

$$\mu_3 = \frac{2e^{\theta}\left(e^{7\theta}+6e^{6\theta}-15e^{5\theta}+2e^{4\theta}+47e^{3\theta}-42e^{2\theta}+15e^{\theta}+10\right)}{\left(e^{2\theta}-e^{\theta}+2\right)^3\left(e^{\theta}-1\right)^3}$$

$$\mu_4 = \frac{2e^{\theta}\left(e^{10\theta}+22e^{9\theta}-24e^{8\theta}-48e^{7\theta}+402e^{6\theta}-548e^{5\theta}+356e^{4\theta}+216e^{3\theta}-307e^{2\theta}+270e^{\theta}+20\right)}{\left(e^{2\theta}-e^{\theta}+2\right)^4\left(e^{\theta}-1\right)^4}$$

The coefficient of variation (C.V), coefficient of skewness $\left(\sqrt{\beta_1}\right)$, coefficient of kurtosis $\left(\beta_2\right)$ and index of dispersion $\left(\gamma\right)$ of DAD are obtained as

$$C.V = \frac{\sigma}{\mu_1'} = \frac{1}{\sqrt{2}}\frac{\sqrt{e^{\theta}\left(e^{4\theta}+2e^{3\theta}-2e^{2\theta}+5\right)}}{\left(e^{2\theta}+e^{\theta}+1\right)}$$

$$\sqrt{\beta_1} = \frac{\mu_3}{\left(\mu_2\right)^{3/2}} = \frac{1}{\sqrt{2e^{\theta}}}\frac{\left(e^{7\theta}+6e^{6\theta}-15e^{5\theta}+2e^{4\theta}+47e^{3\theta}-42e^{2\theta}+15e^{\theta}+10\right)}{\left(e^{4\theta}+2e^{3\theta}-2e^{2\theta}+5\right)^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^{\,2}} = \frac{\left(e^{10\theta} + 22e^{9\theta} - 24e^{8\theta} - 48e^{7\theta} + 402e^{6\theta} - 548e^{5\theta} + 356e^{4\theta} + 216e^{3\theta} - 307e^{2\theta} + 270e^{\theta} + 20\right)}{2e^{\theta}\left(e^{4\theta} + 2e^{3\theta} - 2e^{2\theta} + 5\right)^2}$$

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{e^{\theta}\left(e^{4\theta} + 2e^{3\theta} - 2e^{2\theta} + 5\right)}{\left(e^{2\theta} - e^{\theta} + 2\right)\left(e^{\theta} - 1\right)\left(e^{2\theta} + e^{\theta} + 1\right)}.$$

Table 1 summarizes the nature and behavior of coefficient of variation (C.V), coefficient of skewness, coefficient of kurtosis and index of dispersion (ID) for selected values of the parameter θ.

TABLE 1: Values of descriptive statistics of DAD for varying values of parameter θ.

θ	Values of Descriptive statistics					
	Mean	Variance	C.V	Skewness	Kurtosis	ID
0.5	5.3905	13.2355	0.6749	1.0136	4.6169	2.4553
1.0	1.9381	3.5169	0.9676	1.1411	4.5637	1.8146
1.5	0.8343	1.3082	1.3709	1.6152	5.9718	1.5681
2.0	0.4007	0.5507	1.8519	2.1775	8.5435	1.3745
2.5	0.2091	0.2585	2.4315	2.7875	12.1121	1.2364
3.0	0.1154	0.1323	3.1519	3.4922	17.0323	1.1462
3.5	0.0661	0.0720	4.0603	4.3623	24.2542	1.0895
4.0	0.0387	0.0408	5.2206	5.4749	35.4751	1.0546

It is obvious from above table that the mean, variance, and index of dispersion of DAD are decreasing for increasing values of the parameter θ, while coefficient of variation, coefficient of skewness and coefficient of kurtosis of DAD are increasing for increasing values of parameter θ. Since $\sigma^2 > \mu$, DAD is a suitable model for over-dispersed data.

PARAMETER ESTIMATION

Method of Moment Estimate (MOME): Equating the population mean to the corresponding sample mean, the MOME $\tilde{\theta}$ of the parameter θ is the solution of the following non-linear equation

$$\bar{y}\, e^{3\theta} - 2\left(\bar{y}+1\right)e^{2\theta} + \left(3\bar{y}-2\right)e^{\theta} - 2\left(\bar{y}+1\right) = 0,$$

where $\bar{y}$ is the sample mean.

Maximum Likelihood Estimate (MLE): Let $\left(y_1, y_2, y_3, \ldots, y_n\right)$ be a random sample from DAD (2.3)). The likelihood function, L of (2.4) is given by

$$L = \left(\frac{\left(e^{\theta}-1\right)^3}{e^{\theta}\left(e^{2\theta} - e^{\theta} + 2\right)}\right)^n \prod_{i=1}^{n}\left(1 + y_i^{\,2}\right)e^{-n\theta\bar{y}}$$

and so its natural log likelihood function is thus obtained as

$$\ln L = n\ln\left(\frac{\left(e^{\theta}-1\right)^3}{e^{\theta}\left(e^{2\theta} - e^{\theta} + 2\right)}\right) + \sum_{i=1}^{n}\ln\left(1 + y_i^{\,2}\right) - n\theta\bar{y}$$

The maximum likelihood estimates (MLE) $\hat{\theta}$ of parameter θ is the solution of the non-linear equation

$$\frac{d\ln L}{d\theta}=\frac{3n}{e^{\theta}-1}-\frac{n\left(3e^{3\theta}-2e^{2\theta}+2e^{\theta}\right)}{e^{3\theta}-e^{2\theta}+2e^{\theta}}-n\,\overline{y}=0\,.$$

This gives the following non-linear equation

$$(\overline{y}+3)e^{3\theta}-2(\overline{y}+4)e^{2\theta}+(4\overline{y}+7)e^{\theta}-2(\overline{y}+4)=0\,.$$

DISCUSSION AND GOODNESS OF FIT

Since DAD is over-dispersed, an attempt has been made to fit DAD to over-dispersed frequency distribution and compare its fit with equi-dispersed and over-dispersed discrete distributions. In this section, the goodness of fit of the DAD has been discussed with five count datasets and the fit has been compared with discrete Shanker distribution (DSD), discrete Lindley distribution (DLD), Poisson distribution (PD), Poisson-Lindley distribution (PLD) introduced by Sankaran (1970) and Poisson-Akash distribution (PAD) introduced by Shanker (2017).[14,17] The dataset in table 2 is the data relating to the production of chromosome structural changes in Tradescannia microspores in relation to dosages, intensity and temperature, available in Catcheside *et al* (1946).[18] The dataset in tables 3, 4 and 5 are the number of days that experienced X thunderstorm events at Cape kennedy, Florida for the 11-year period of record in June, July and the summer, January 1957 to December 1967, available in Falls *et al* (1971) and Carter (2001).[19,20] The dataset in table 6 is the data relating to the number of red mites on apple leaves, available in Gosset (1908).[21]

TABLE 2: Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by Streptonigen (NSC-45383), Exposure-70 μg |kg.

Class/Exposure	Observed	Expected Frequncy					
μg|kg	Frequency	DAD	DSD	DLD	PD	PLD	PAD
0	200	193.2	187.9	184.1	172.5	191.8	194.1
1	57	65.7	75.7	79.8	95.4	70.3	67.6
2	30	27.9	25.4	25.9	26.4	24.9	24.5
3	7	9.5	7.8	7.5	4.9	8.6	8.9
4	4	2.8	2.3	2.0	0.7	2.9	3.2
5	0	0.7	0.6	0.5	0.1	1.0	1.1
6	2	0.2	0.3	0.2	0.0	0.5	0.6
Total	300	300.0	300.0	300.0	300.0	300.0	300.0
ML estimate $(\hat{\theta})$		1.77177	1.43696	1.52919	0.55333	2.35334	2.62674
χ^2		1.55	6.60	9.31	29.68	3.91	3.12
d.f.		2	2	2	2	2	2
p-value		0.6702	0.0860	0.0255	0.0000	0.2712	0.3731

TABLE 3: Frequencies of the observed number of days that experienced X thunderstorm events at Cape kennedy, Florida for the 11-year period of record in the month of June, January 1957 to December 1967.

X	Observed	Expected Frequency					
	Frequency	DAD	DSD	DLD	PD	PLD	PAD
0	187	186.2	177.8	174.3	155.6	185.3	187.9
1	77	77.7	92.0	95.2	117.0	83.5	80.3
2	40	40.5	38.2	39.0	43.9	36.0	35.3
3	17	16.9	14.3	14.2	11.0	15.0	15.4

4	6	6.0	5.1	4.9	2.1	6.1	6.6
5	2	2.0	1.7	1.6	0.3	2.5	2.7
6	1	0.7	0.9	0.8	0.1	1.6	1.8
Total	330	330.0	330.0	330.0	330.0	330.0	330.0
ML estimate $(\hat{\theta})$		1.5673	1.2478	1.29773	0.751515	1.80423	2.139736
χ^2		0.03	3.86	5.38	31.93	1.37	1.33
d.f.		3	3	3	2	3	3
p-value		0.9986	0.4250	0.2506	0.0000	0.8487	0.8564

TABLE 4: Frequencies of the observed number of days that experienced X thunderstorm events at Cape kennedy, Florida for the 11-year period of record in the month of July, January 1957 to December 1967.

X	Observed	Expected Frequency					
	Frequency	DAD	DSD	DLD	PD	PLD	PAD
0	177	177.9	167.9	165.1	142.3	177.7	180.0
1	80	81.8	98.1	100.4	124.4	88.0	84.8
2	47	47.0	45.0	45.8	54.3	41.5	40.9
3	26	21.6	18.6	18.6	15.8	18.9	19.4
4	9	8.4	7.2	7.1	3.5	8.4	8.9
5	2	4.3	4.2	4.0	0.7	6.5	7.0
Total	341	341.0	341.0	341.00	341.0	341.0	341.0
ML estimate $(\hat{\theta})$		1.47054	1.15984	1.19041	0.873900	1.583536	1.938989
χ^2		1.17	6.88	7.98	39.74	5.15	4.99
d.f.		3	3	3	2	3	3
p-value		0.8833	0.1424	0.0923	0.0000	0.2725	0.2886

TABLE 5: Frequencies of the observed number of days that experienced X thunderstorm events at Cape kennedy, Florida for the 11-year period of record in the summer, January 1957 to December 1967.

X	Observed	Expected Frequency					
	Frequency	DAD	DSD	DLD	PD	PLD	PAD
0	549	549.1	521.6	511.8	449.2	547.5	555.1
1	246	240.9	287.1	295.7	364.9	259.0	249.2
2	117	132.1	125.4	128.1	148.2	116.9	144.9
3	67	57.9	49.4	49.3	40.1	51.2	52.3
4	25	21.6	18.4	17.8	8.1	21.9	23.2
5	7	7.2	6.6	6.2	1.3	9.2	10.0
6	1	3.2	3.5	3.1	0.2	6.3	7.3
Total	1012	1012.0	1012.0	1012.0	1012.0	1012.0	1012.0
ML estimate $(\hat{\theta})$		1.517111	1.202505	1.24195	0.812253	1.687436	2.026172
χ^2		4.35	16.96	21.47	142.57	9.60	14.75
d.f.		4	4	4	3	4	4
p-value		0.4998	0.0046	0.0007	0.0000	0.0874	0.0115

TABLE 6: Observed and expected number of red mites on apple leaves, available in Gosset (1908).

Number of red mites	Observed	Expected Frequency					
	Frequency	DAD	DSD	DLD	PD	PLD	PAD
0	70	66.1	61.7	60.6	47.6	67.3	68.3
1	38	35.9	44.1	44.2	54.6	38.9	37.7
2	17	24.4	23.8	24.1	31.3	21.2	21.0
3	10	13.3	11.4	11.7	12.0	11.2	11.4
4	9	6.1	5.2	5.3	3.4	5.7	5.9
5	3	2.5	2.2	2.3	0.8	2.9	3.0
6	2	1.0	0.9	1.0	0.2	1.4	1.5

7	1	0.7	0.7	0.8	0.1	1.4	1.2
Total	150	150.0	150.0	150.0	150.0	150.0	150.0
ML Estimate $(\hat{\theta})$		1.302757	1.019163	1.0095	1.14666	1.260163	1.628089
χ^2		5.56	8.08	8.00	26.50	2.26	1.98
d.f.		3	3	3	2	3	3
p-value		0.2344	0.0889	0.0915	0.0000	0.6942	0.7403

It is obvious from the goodness of fit in Tables 2, 3, 4, 5, and 6 that except in table 6, DAD gives much closer fit than other considered discrete distributions. Hence, DAD can be considered an important discrete distribution over these distributions. In Table 6, PAD gives much closer fit than other discrete distributions.

CONCLUDING REMARKS

In this paper, a discrete Akash distribution (DAD), a discrete analogue of continuous Akash distribution, has been proposed and investigated. Its moment generating function, moments and moments based measures including coefficients of variation, skewness, kurtosis, index of dispersion have been obtained and their nature has been discussed numerically. Both the method of moments and the method of maximum likelihood estimation have been discussed for estimating its parameter. The goodness of fit of DAD has been explained using some real datasets. The DAD gives much closer fit over PD, PLD, PAD, DLD, and DSD in majority of datasets and hence it can be considered an important discrete distribution over these discrete distributions.

Acknowledgement

Authors are grateful to the Editor-In-Chief of the journal and three independent anonymous reviewers for their constructive comments which improved the quality and the presentation of the paper.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

***Idea/Concept:** Rama Shanker; **Design:** Rama Shanker; **Control/Supervision:** Rama Shanker; **Data Collection and/or Processing:** Berhane Abebe; **Analysis and/or Interpretation:** Berhane Abebe; **Literature Review:** Berhane Abebe; **Writing the Article:** Berhane Abebe; **Critical Review:** Rama Shanker; **References and Fundings:** Rama Shanker; **Materials:** Rama Shanker.*

REFERENCES

1. Lai CD. Issues concerning constructions of discrete lifetime models. Qual Techno Quant Mang 2013;10(2):251-62.
2. Good LJ. The population frequencies of species and the estimation of population parameters. Biometrika 1953;40(3-4):237-64.
3. Kulasekera KB, Tonkyn DW. A new discrete distribution, with application to survival, dispersal and dispersion. Commun Stat Simul Comput 1992;21(2):499-518.
4. Doray LG, Luong A. Efficient estimators for the good family. Commun Stat Simul Comput 1997;21:499-518.

5. Sato H, Ikota M, Aritoshi S, Masuda H. A new defect distribution in meteorology with a consistent discrete exponential formula and its applications. IEEE Trans Semicond Manufactur 1999;12(4):409-18.
6. Nekoukhou VM, Alamatsaz MH, Bidram H. A discrete analog of the generalized exponential distribution. Communications in Statistics-Theory & Methods 2012;41(11):2000-13.
7. Josmar M, Wesley BDS, Ricardo PO. On the discrete shanker distribution, submitted to Chilean. Journal of Statistics 2017.
8. Shanker R. Shanker distribution and its applications. Int J Stat Appl 2015a;5(6):338-48.
9. Berhane A, Shanker R. A discrete Lindley distribution and its Applications, to appear in. Biom Biostat Int J 2018.
10. Lindley DV. Fiducial distributions and Bayes' theorem. J R Stat Soc Series B Stat Methodol 1958;20(1):102-7.
11. Nakagawa T, Osaki S. The discrete Weibull distribution. IEEE Trans Reliability 1975;R-24(5):300-1.
12. Kemp AW. Classes of discrete lifetime distributions. Communications in Statistics-Theory & Methods 2004;33(12):3069-93.
13. Shanker R. Akash distribution and its applications. International Journal of Probability and Statistics 2015b;4(3):65-75.
14. Shanker R. The discrete poisson-akash distribution. International Journal of Probability and Statistics 2017;6(1):1-10.
15. Keilson J, Gerber H. Some results for discrete unimodality. J Am Stat Assoc 1971;66(334):386-9.
16. Grandell J. Mixed Poisson Processes. 1st ed. London: Chapmann & Hall; 1997. p.268.
17. Sankaran M. The discrete Poisson-Lindley distribution. Biometrics 1970;26(1):145-9.
18. Catcheside DG, Lea DE, Thoday JM. The production of chromosome structural changes in Tradescannia microspores in relation to dosages, intensity and temperature. J Genet 1946;47:137-49.
19. Falls LW, Wilford WO, Carter MC. Probability distribution for thunderstorm activity at Cape Kennedy, Florida. Journal of Applied, Meteorology 1971;10:97-104.
20. Carter MC. A model for thunderstorm activity: use of the compound negative binomial-positive binomial distribution. Journal of the Royal Statistical Society, Series C (Applied Statistics) 2001;2:196-201.
21. Gosset WS. The probable error of the mean. Biometrika 1908;6(1):1-25.